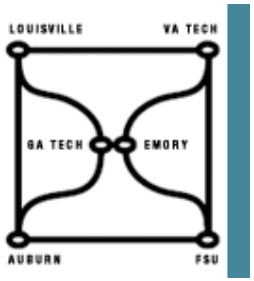
A Practical, Working and Replicable Approach to ETD Preservation

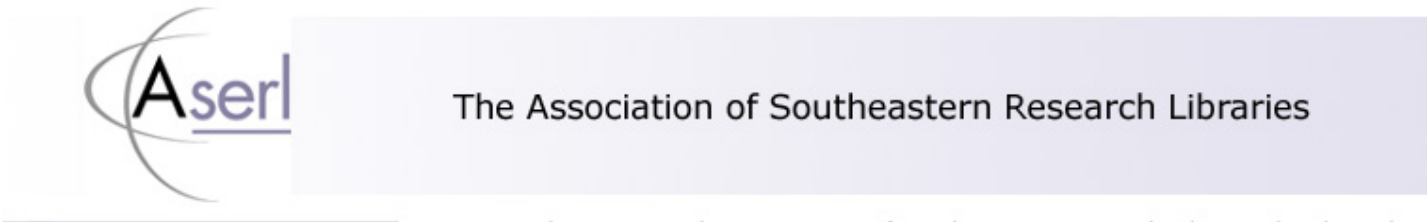
Catherine M. Jannik, Georgia Institute of Technology
Robert H. McDonald, Florida State University
Gail McMillan, Virginia Polytechnic Institute and State University

8th International Symposium on Electronic Theses and Dissertations, Sept. 29, 2005



http://www.metaarchive.org/

- The MetaArchive Project is a collaborative venture of Emory University, Georgia Tech, Virginia Tech, Florida State University, Auburn University, University of Louisville, and the Library of Congress. The project is part of the National Digital Information Infrastructure and Preservation Program (NDIIPP) supported by the Library of Congress.



- The ASERL-LOCKSS-ETD Initiative is a joint project between the LOCKSS (Lots of Copies Keep Stuff Safe) Program at Stanford University and the University Libraries of the Florida State University, Georgia Institute of Technology, University of Kentucky, University of Tennessee, Vanderbilt University, and the Virginia Polytechnic Institute. http://www.aserl.org/

http://www.lockss.org

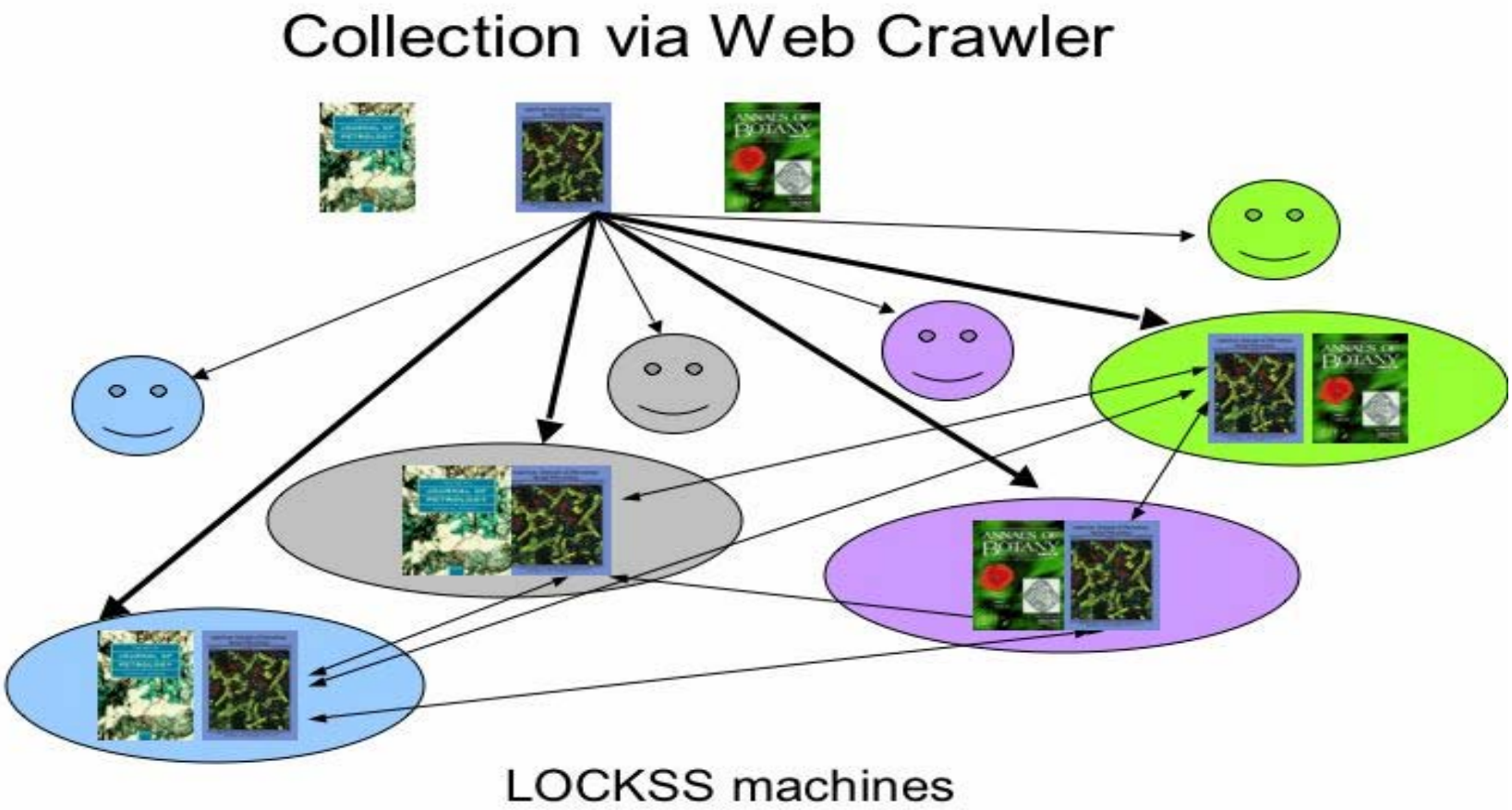
- "...let us save what remains: not by vaults and locks which fence them from the public eye and use in consigning them to the waste of time, but by such a multiplication of copies, as shall place them beyond the reach of accident."

  Thomas Jefferson, 1791

Distributed Archiving Strategies

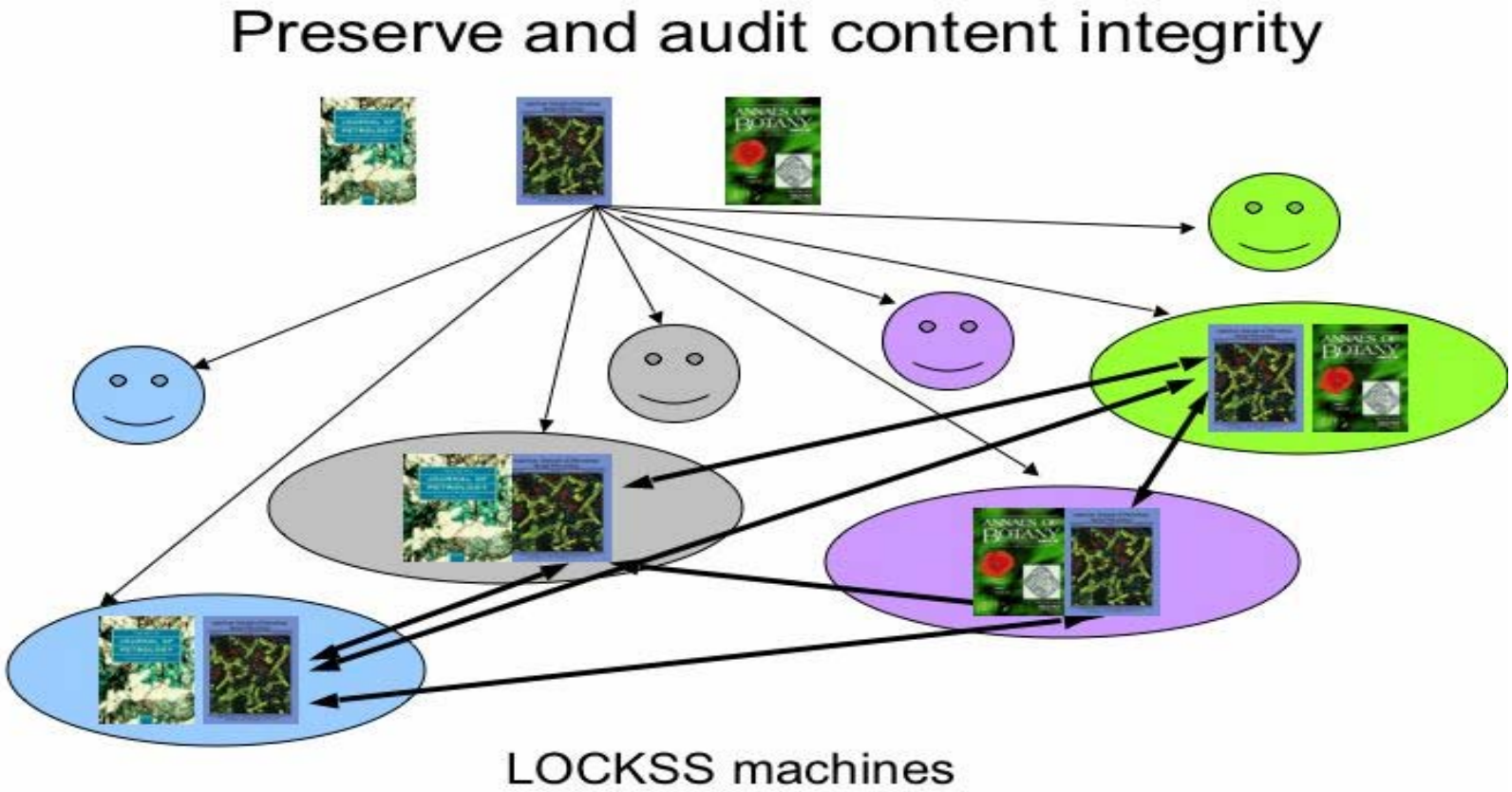
- LOCKSS for Electronic Journals
- LOCKSS for ETDs

LOCKSS for Electronic Journals



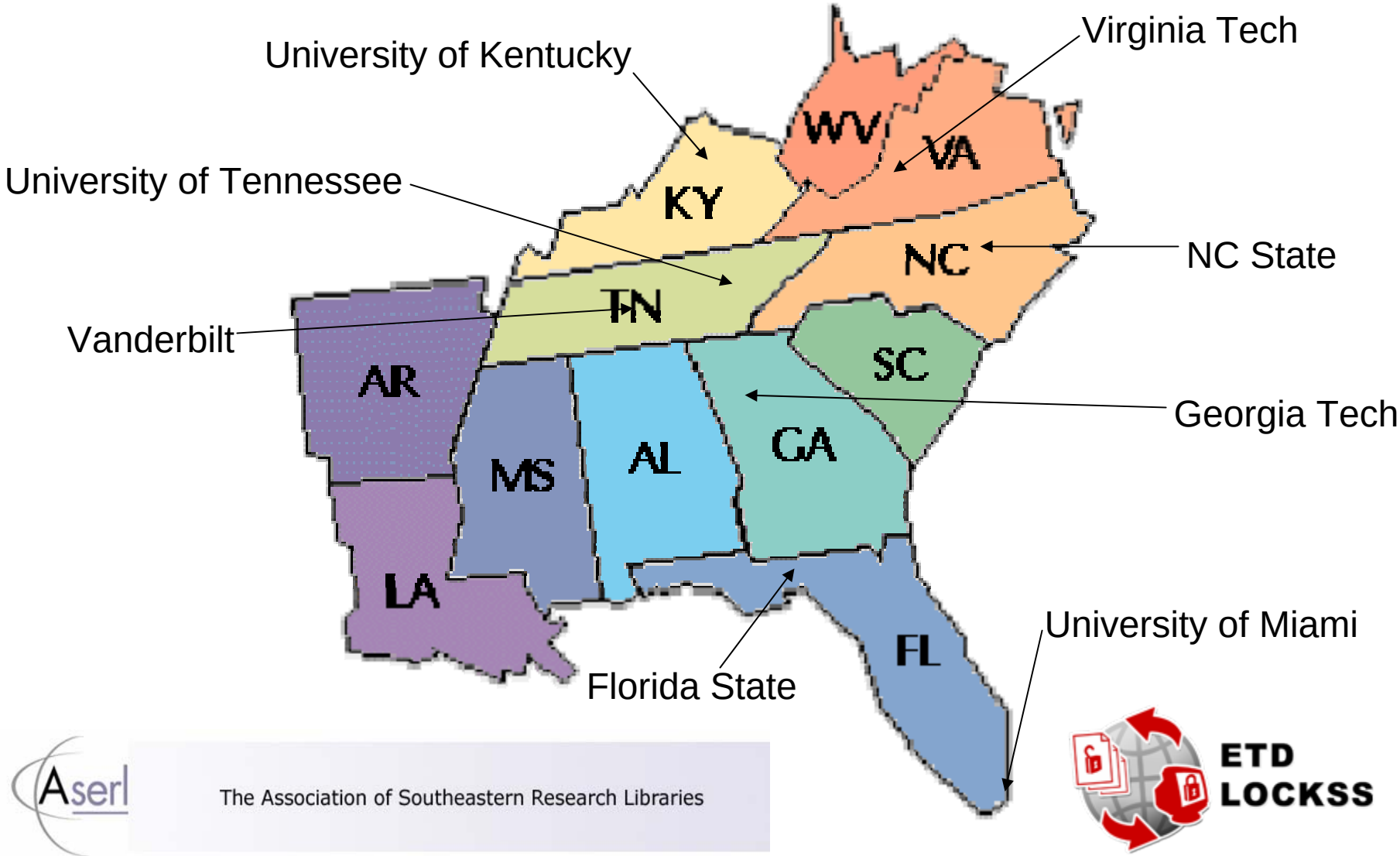
http://lockss.stanford.edu/works/how_it_works.htm

LOCKSS for Electronic Journals



http://lockss.stanford.edu/works/how_it_works.htm

LOCKSS for ETDs

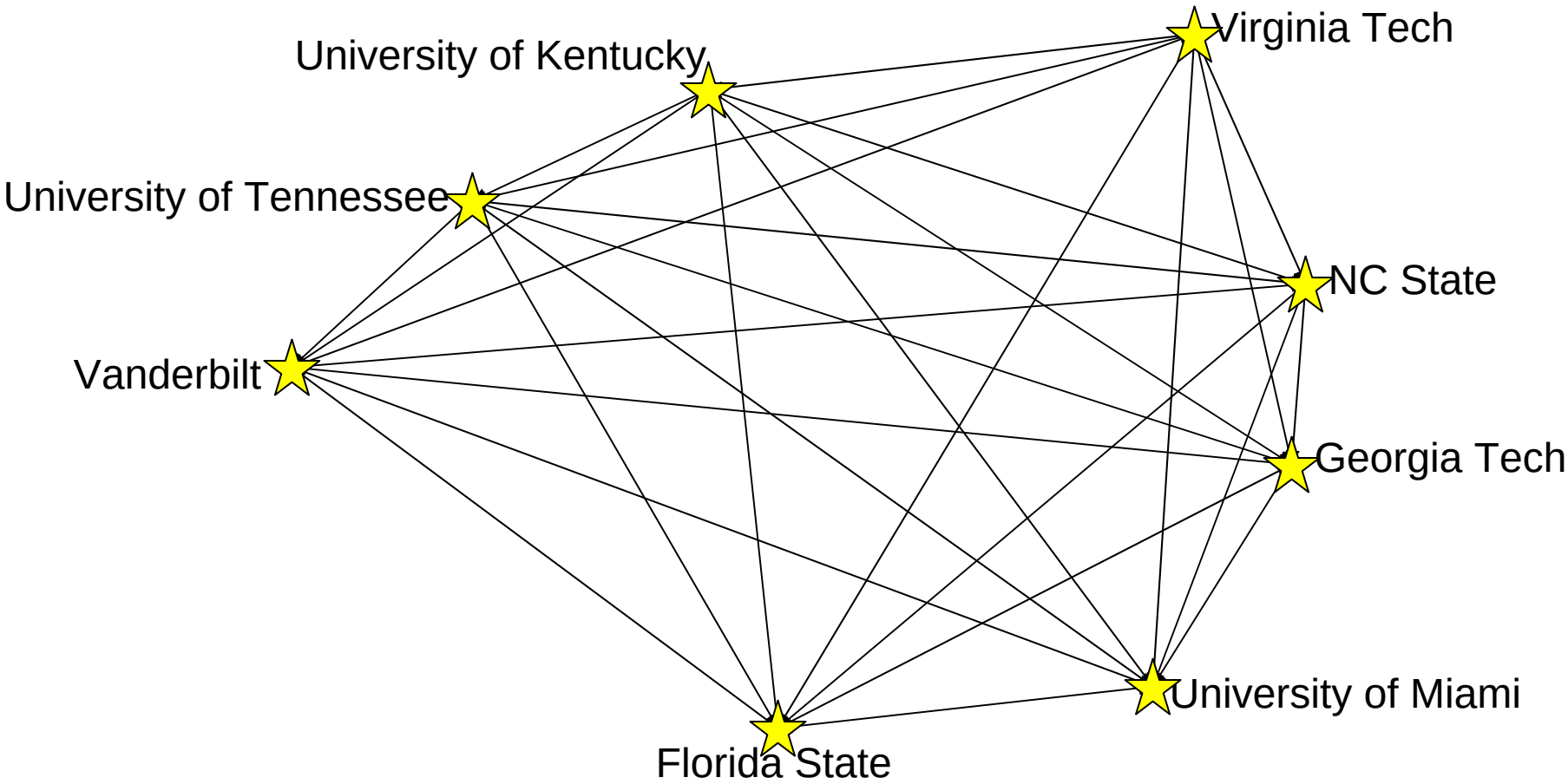


Aserl

The Association of Southeastern Research Libraries

ETD LOCKSS

LOCKSS for ETDs



Aserl

The Association of Southeastern Research Libraries

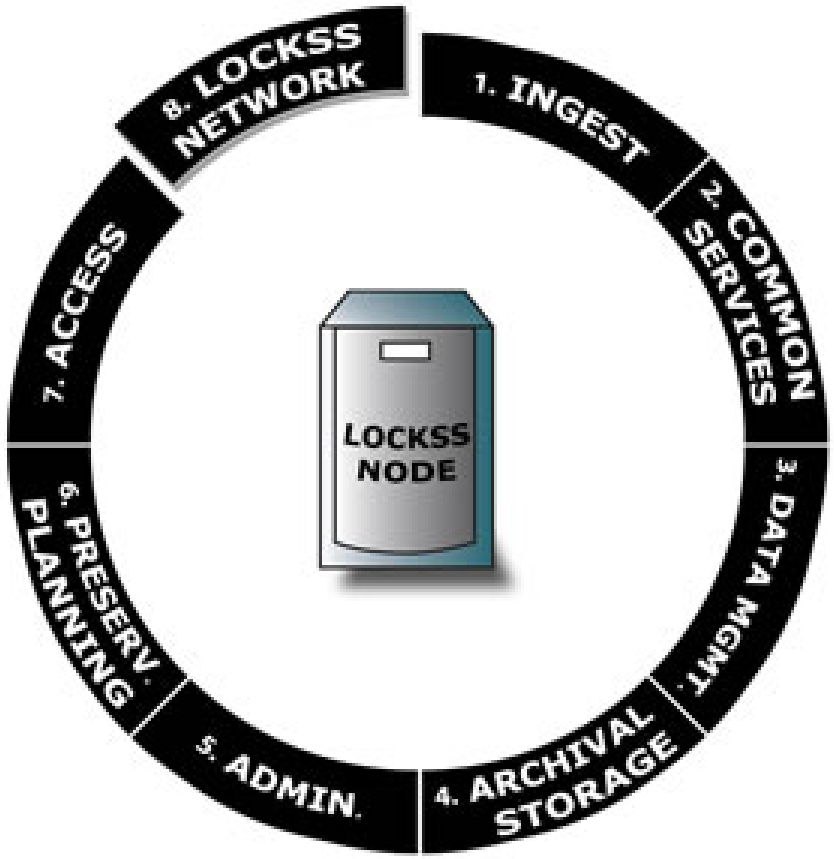
ETD LOCKSS

Technical Infrastructure

- Overall Goals of TI
  - Build on successful LOCKSS open-source model
  - Create dark archive for locally produced digital content
  - Use *off-the-shelf* hardware
  - Use open-source software (currently all)
  - Create ease of replication
  - Demonstrate LOCKSS scalability
  - Enable benefits of Internet2 network

ETD LOCKSS

LOCKSS and the OAIS Framework



ETD LOCKSS

Hardware

- Off-the-Shelf Strategy
  - Dell/Intel Based Hardware
    - Could easily be HP or SUN Intel Based Hardware etc.
    - Could be old or new desktops w/large hard drives.
  - New Low Cost SATA SAN
    - EMC AX100
      - $4.00 per GB (already dropping in price)

ETD LOCKSS

Software

- Operating System
  - RedHat Enterprise Linux AS v. 4
    - Ease of update management and experience w/OS
      - Setup can easily be set up on other versions of Linux using kickstart configuration.
    - JAVA SDK
  - Also tested with CentOS Linux Distribution
- LOCKSS Content Ingestion/Replication
  - LOCKSS Daemon 1.10.5 – 6-8 week updates w/RPM files produced by LOCKSS.
- Conspectus Database
  - MySQL/PHP Web Interface – Integrated with LOCKSS Plugin Registry – Viewable to all - Editable by Members
    - http://metascholar4.library.emory.edu/coll_desc/final/
- MetaArchive Collection Description Metadata Schema

Standards

- OAIS Reference Model
  - LOCKSS Compliance (Pull Methodology)
- Multiple Submission Information Package (SIP) Model
  - OAI-PMH 2.0
    - Using as alternative to current LOCKSS AU strategy
  - LOCKSS Audit Procedure
- Modified UKOLN RSLP Collection Description
  - Basis for MetaArchive Collection Conspectus
    - http://www.metaarchive.org/pdfs/conspectus_md_2005.html

ETD LOCKSS

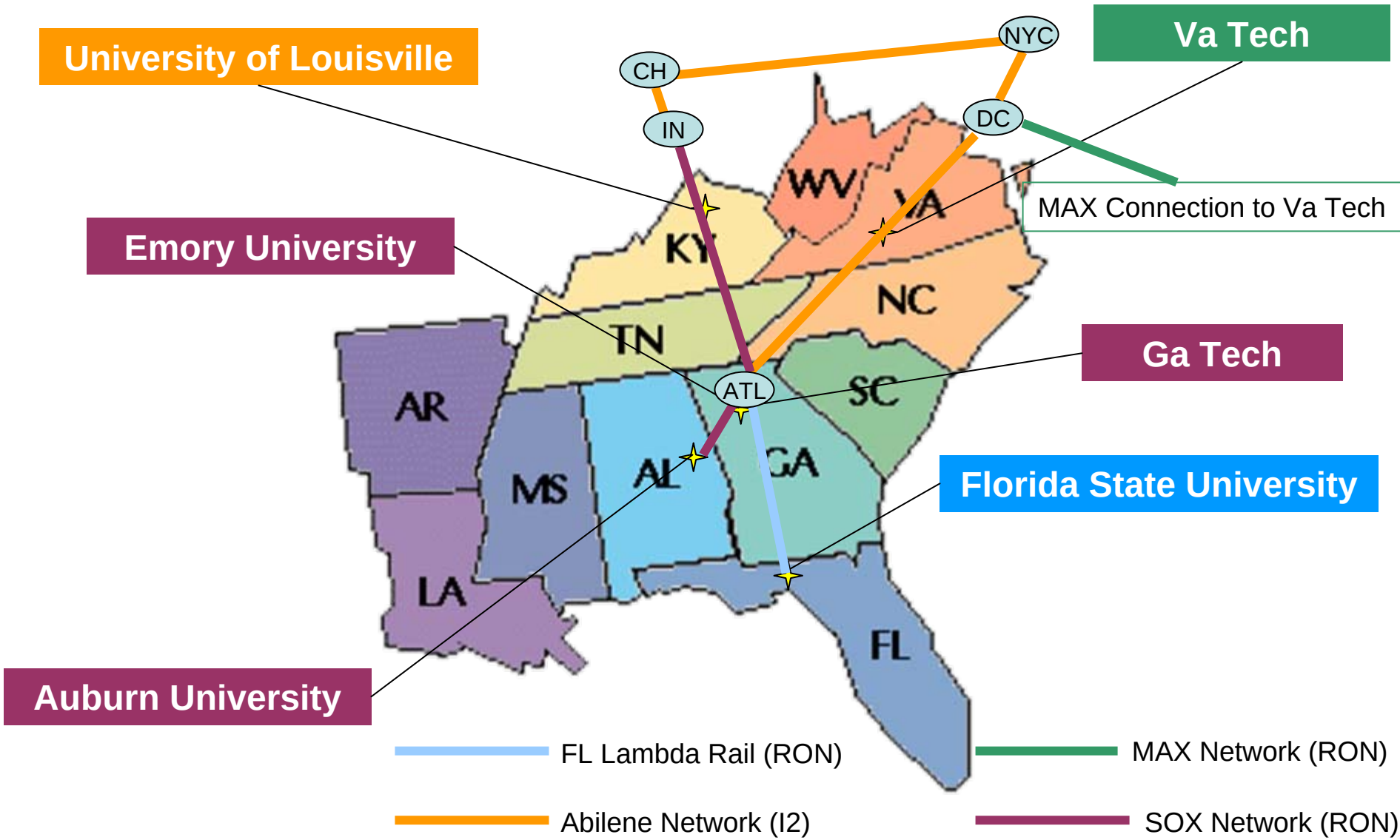
EXAMPLE SETUP

- Enterprise (3TB)
  - Dell PowerEdge Server 1850 LOCKSS - $3500
  - Dell PowerEdge Server 1850 Firewall - $2500
  - Dell/EMC AX100 SAN (3TB) - $10,000
  - RedHat Enterprise AS – 2@$50 = $100
  - UPS - $700
  - Server Rack - $1200
- **Grand Total - $16,800.00**
  - w/ Rack - $18,000.00

- Desktop (200Gb)
  - Intel Based Desktop LOCKSS (200Gb) - $500
  - Intel Based Desktop Firewall - $350
  - CentOS Linux - $0
  - UPS - $50
- **Grand Total - $900.00**

ETD LOCKSS

MetaArchive Network via Internet2



NDLTD and LOCKSS

- NDLTD Board of Directors
- Conclusion and Call for Participation
- Why not an Electronic Thesis and Dissertation International Preservation Network (ETD-IPN)?

Further Reading

- Consultative Committee for Space Data Systems (CCSDS). (2002). *Reference Model for an Open Archival Information System (OAIS), Blue Book, Issue 1, January 2002, ISO 14721:2003* http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf

- Rosenthal et. al. *Requirements for Digital Preservation Systems: A Bottum-Up Approach* http://xxx.arxiv.cornell.edu/abs/cs.DL/0509018

- RLG-OCLC. (2002). *Trusted Digital Repositories: Attributes and Responsibilities. Mountain View, CA.* http://www.rlg.org/legacy/longterm/repositories.pdf

ETD LOCKSS